

# Ingeniería de Aplicaciones para la Web Semántica

## Clase 05

*Intro. a la web semántica*

Mg. A. G. Stankevicius

Segundo Cuatrimestre

2005





# Copyright

- Copyright © 2005 A. G. Stankevicius.
- Se asegura la libertad para copiar, distribuir y modificar este documento de acuerdo a los términos de la GNU Free Documentation License, Version 1.2 o cualquiera posterior publicada por la Free Software Foundation, sin secciones invariantes ni textos de cubierta delantera o trasera.
- Una copia de esta licencia está siempre disponible en la página <http://www.gnu.org/copyleft/fdl.html>.
- La versión transparente de este documento puede ser obtenida en <http://cs.uns.edu.ar/~ags/IAWS>.



# Contenidos

- La web de hoy en día.
- Inconvenientes de la web sintáctica.
- La propuesta de la web semántica.
- El impacto de la web semántica.
- Tecnologías asociadas
- El rol de la metainformación y de las ontologías asociadas.
- La lógica y la obtención de inferencias.
- Los agentes de software



# La web de hoy en día

- La mayor parte del contenido de la web esta **orientado al consumo humano**:
  - ➔ Carece de información adicional respecto de la estructura de la información contenida.
  - ➔ Aún en el caso que el contenido presentado provenga de bases de datos, la información presentanda al usuario deja de lado las consideraciones acerca de su estructura.



# Usos típicos de la web

- La gente hace un uso bastante estandarizado de la web:
  - ➔ Busca y hace uso de variada información.
  - ➔ Se pone en contacto con otras personas, empleando a la web como medio para conocer o acceder a las mismas.
  - ➔ Repasa y compara catálogos de sitios de venta de productos online.
  - ➔ Hace compras en esos sitios completando formularios HTML.



# Motores de búsqueda

- No existen programas que asistan al usuario típico en las tareas que realiza con la web
- Notable excepción: los motores de búsqueda basados en palabras claves.
- La web no sería lo que es hoy en día sin contar con los motores de búsqueda.



# Problemas de la búsqueda basada en palabras claves

- Una consulta, miles de respuestas.
- Una consulta, cero respuestas.
- Los resultados son dependientes de la palabras clave empleadas.
- Los resultados son páginas web.
- Sin un humano interpretando los datos, las búsquedas resultan inconducentes.
- Los resultados no son fáciles de integrar en otras herramientas.



# La web sintáctica

- El problema central de la web sintáctica de hoy en día es la **ausencia de información semántica**.
- Frases sintacticamente similares pueden tener significados diferentes, o incluso contrapuestos:
  - ➔ La segunda presidencia de Mendez fue muy buena.
  - ➔ La segunda presidencia de Mendez fue muy buena, destruyó la industria nacional.



# Una posible solución...

- La web semántica propone representar el contenido de la web de forma que sea procesable por las computadoras de manera relativamente simple.
- La idea es hacer uso de diversas técnicas para sacar provecho de esta representación.
- La web semántica no compite con la web sintáctica, por el contrario, surgirá a partir de la web de hoy en día.



# El impacto de la web semántica

- La web semántica posiblemente revolucione la forma en que hacemos uso del conocimiento.
- La mayor parte de la información con la que contamos hoy en día viene en formatos poco estructurados:
  - ➔ Texto.
  - ➔ Audio.
  - ➔ Videos.



# Limitaciones en la gestión actual del conocimiento

- **Búsqueda de la información.**
  - ➔ Búsqueda basadas en palabras clave.
- **Extracción de la información.**
  - ➔ Sólo los humanos están en condiciones de encontrar, categorizar, interpretar o recombinar la información.
- **Mantenimiento de la información.**
  - ➔ Eliminación de inconsistencias, etc.
- **Gerenciación de la información.**
  - ➔ Especificación de restricciones de acceso.



# Gestión del conocimiento en el marco de la web semántica

- La información será organizada en función de su significado.
- Se contará con herramientas automatizadas para descubrir información y mantenerla actualizada.
- Se podrán efectuar y responder consultas de naturaleza semántica, sobre la base de uno o más documentos.
- Será posible especificar quién o quienes pueden tener acceso a la información.



# El comercio electrónico y la web semántica

- Un escenario típico:
  - ➔ Un usuario visita varios sitios, repasa sus ofertas, selecciona qué comprar y de contar con los medios efectua la compra.
- Siguiendo los consejos de Lita, habría que visitar todos o casi todos los sitios:
  - ➔ El tiempo perdido termina costando más que el dinero ahorrado.
- Los **shopbots** pueden simplificar parte de esta tarea.



# Limitaciones de los shopbots

- Se centran exclusivamente en **wrappers**.
- Estos wrappers tienen que ser adaptados cada vez que el sitio web en cuestión cambie de apariencia.
- Los wrappers extraen información a partir de un análisis textual de los datos.
  - ➔ Son propensos a los errores.
  - ➔ Limita el tipo de información a la cual pueden acceder.



# Comercio electrónico en el marco de la web semántica

- Se contará con agentes de software capaces de interpretar la información asociada a los productos así como los términos del servicio contratado.
- Se podrá tener en cuenta la información en relación a la reputación del sitio web.
- Los agentes más sofisticados podrán incluso conducir negociaciones o deliberaciones automatizadas.



# Los servicios web y la web semántica

- El paradigma “*software como servicio*” tiene el potencial de generar un gran mercado.
- Actualmente existe la infraestructura necesaria para sustentar servicios web de diversa índole.
- La incorporación de información semántica a la web extiende el horizonte de posibilidades para los servicios web.



# Tecnologías asociadas a la web semántica

- La web semántica se centra en torno a un conjunto reducido de tecnologías:
  - ➔ La **metainformación**.
  - ➔ Las **ontologías**.
  - ➔ La **lógica** y la obtención de **inferencias**.
  - ➔ Los **agentes de software**.
- En lo que resta repasaremos el rol específico de cada una de estas tecnologías.



# La metainformación y su actual carencia

- El contenido de la web actual está preparado pensando en los humanos
  - ➔ Énfasis en **cómo mostrar** y no en **qué es**.
- La información contenida en la web está codificado mayoritariamente en el lenguaje HTML.
- Las páginas web contiene bastante poca metainformación, apenas un conjunto de palabras clave para que indexen los motores de búsqueda.



# La falta de metainformación

- Consideremos el siguiente fragmento de código HTML:

```
<h1>Curso de Cocina</h1>
```

```
Bienvenido a la página web del curso de  
cocina de Doña Petrona y Yiya Murano.
```

```
<h2>Horario de clase</h2>
```

```
Ma 10am-12am<br>
```

```
Jue 8am-12am<br>
```

```
<p>Atención, la semana del CACIC no  
habrá clases en toda la semana!</p>
```



# El problema de la falta de metainformación

- Los humanos posiblemente no tengan problema entendiendo la información
- Pero... ¿y los agentes de software?
  - ➔ ¿Cómo determinar cuántas personas están a cargo del curso?
  - ➔ ¿Cómo determinar en qué días y a qué hora hay clase?
  - ➔ ¿Cómo saber qué semana no habrá clase?
  - ➔ ¿Cómo caer en cuenta que el curso es un scam?



# Una representación alternativa

- El fragmento anterior puede ser reescrito de una manera más práctica:

```
<course>
  <name>Curso de Cocina</name>
  <topic>Cocina</topic>
  <teachers>
    <teacher>Doña Petrona</teacher>
    <teacher>Yiya Murano</teacher>
  </teachers>
  . . .
</course>
```



# Explicitación de la metainformación

- La representación alternativa propuesta puede ser procesada de forma mucho más simple por las computadoras.
- Cuenta con más **metainformación**, información acerca de la información.
- La web semántica no depende de manipular información textual sino de hacer un uso adecuado de la metainformación presente.



# Las ontologías

- Este término surge del ámbito de la filosofía:
  - ➔ Significa literalmente *“el estudio de la naturaleza de la existencia”*.
- En ciencias de la computación le damos otro significado:
  - ➔ Una ontología es una especificación **explícita** y **formal** de una **conceptualización**.



# Componentes de una ontología

- Las ontologías están compuestas de dos aspectos esenciales:
  - **Términos**, denotando los conceptos importantes (e.g., las clases de objetos) presentes en el dominio en cuestión.
  - **Relaciones** entre estos términos, usualmente la jerarquía de clases existente entre los mismos.



# Otros componentes presentes en una ontología

- Propiedades.
  - ➔ Doña Petrona dicta el curso Cocina 101.
- Restricciones en los valores.
  - ➔ Sólo los profesores pueden dictar cursos.
- Explicitación de independencia.
  - ➔ Los conjuntos de profesores y de alumnos son disjuntos.
- Relaciones lógicas entre objetos.
  - ➔ Todo curso debe tener al menos un profesor y al menos cinco alumnos.



# El rol de las ontologías en la web sintáctica

- Las ontologías proveen una **base compartida de entendimiento** en relación a un cierto dominio:
  - ➔ Permiten superar diferencias en la terminología.
  - ➔ Establecer mapeos entre diferentes ontologías.
- Muchos sitios web ayudan a la navegación mediante **crumbs**, de estructura netamente jerárquica.



# El rol de las ontologías en las búsquedas en la web

- Las ontologías permiten **mejorar la correctitud** de los resultados provistos:
  - ➔ Al indexar en base a los conceptos definidos en una ontología se pueden evitar los resultados espurios.
- Los motores de búsquedas puede apelar a generalizaciones o particularizaciones:
  - ➔ Si se encuentra mucho resultados, se pueden sugerir términos más específicos.
  - ➔ Si no se encuentran resultados, se puede probar con palabras clave más genéricas.



# Lenguajes de representación de ontologías para la web

- KIF (Knowledge Interchange Format).
- RDF (Resource Description Framework) y RDFS (RDF Schema).
- DAML (DARPA Agent Markup Language).
- OIL (Ontology Inference Layer).
- DAML+OIL, esencialmente la combinación de los anteriores.
- OWL (Web Ontology Language).



# La lógica y la obtención de inferencias

- La lógica es la disciplina que estudia los **patrones del razonamiento**.
- Es a la ciencia de la computación lo que la matemática es a la física.
- Emplea **lenguajes formales** para **expresar conocimiento**.
- Cuenta con una **semántica clara**.
- A partir del conocimiento disponible es posible **computar** de forma **automática** las conclusiones que de él se derivan.



# Ejemplo de inferencia

- Considerando el siguiente conocimiento:
  - $\text{teacher}(X) \rightarrow \text{teachers}(X)$
  - $\text{teachers}(X) \rightarrow \text{staff}(X)$
  - $\text{teacher}(\text{Yiya})$
- ...es posible inferir que:
  - $\text{teachers}(\text{Yiya})$
  - $\text{staff}(\text{Yiya})$
  - $\text{teacher}(X) \rightarrow \text{staff}(X)$



# La lógica vs. las ontologías

- Las inferencias anteriores explicitaron conocimiento que posiblemente estaba presente en la ontología asociada:
  - La lógica permite explicitar conocimiento que posiblemente figuraba de forma implícita en la especificación de la ontología.
  - Por otra parte, puede también explicitar inconsistencias o relaciones no tan evidentes entre los objetos.
- La lógica es más general que las ontologías.



# El compromiso de siempre

- Cuanto más expresiva es una lógica, más costoso resulta obtener nuevas inferencias.
  - Más costoso puede llegar a ser “imposible”.
- El ejemplo anterior involucraba una cantidad finita de reglas **si-entonces**, sobre un dominio a su vez finito.
  - Este subconjunto de la lógica es tratable.
  - Existen algoritmos muy eficientes para razonar en este contexto.



# Explicando las inferencias

- Es posible enumerar los pasos que fueron necesarios para alcanzar una cierta conclusión.
  - ➔ Esta secuencia constituye una **explicación** de por qué se llegó a esa conclusión.
- Las explicaciones pueden incrementar la confianza de los usuario típicos en la web semántica:
  - ➔ Por caso, el botón de “**Oh yeah?**”, propuesto por el mismo Tim Berners-Lee.

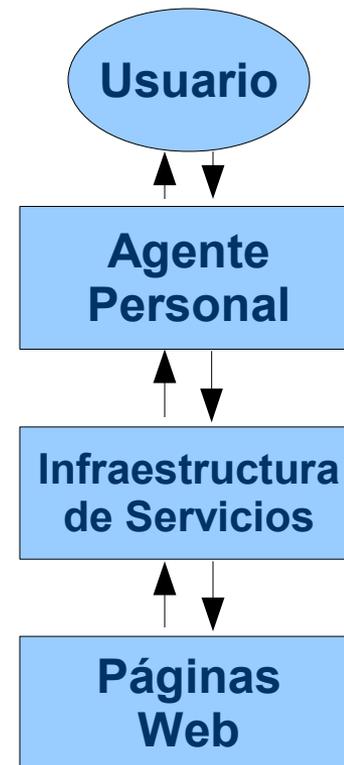
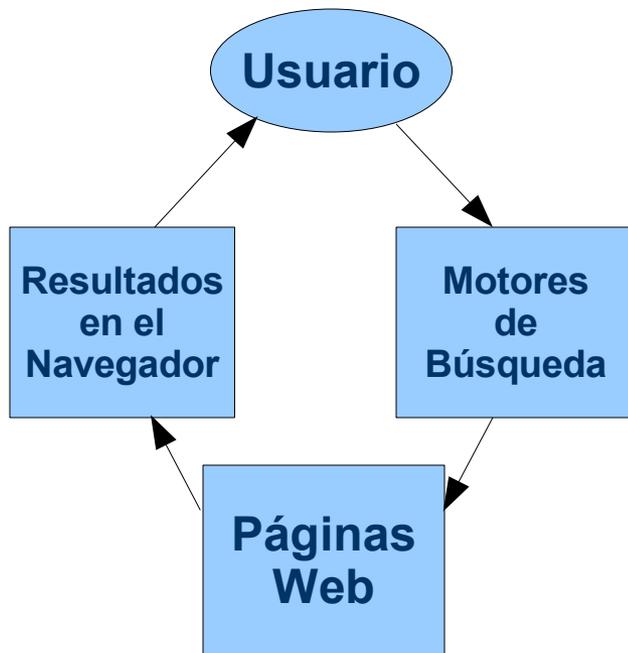


# Agentes de software

- Los agentes de software son **entidades** que **actúan** de manera **autónoma** y **proactiva**.
- Un agente de software en la web semántica podría:
  - Recibir tareas delegadas por el usuario.
  - Buscar información en la web.
  - Comunicarse con otros agentes.
  - Tomar decisiones por el usuario.
  - Retornar con los resultados deseados.



# Agentes personales inteligentes





# La web semántica y los agentes de software

- **Metainformación**
  - ➔ Identificar y extraer información de la web.
- **Ontologías**
  - ➔ Realizar búsquedas efectivas, pudiendo interpretar los resultados obtenidos.
  - ➔ Comunicarse con otros agentes.
- **Lógica y obtención de inferencias**
  - ➔ Procesar la información recuperada, pudiendo obtener nuevas inferencias en base a ese conocimiento.



# Simbiosis entre la web semántica y los agentes

- Toda innovación en las tecnologías asociadas a ambos campos redundará en avances para las dos disciplinas:
  - Lenguajes de comunicación entre agentes.
  - Representación de conocimiento.
  - Representación de otras modalidades (deseos, intenciones, etc.).
  - Creación y mantenimiento de modelos de otras entidades.